

Clinical Characteristics Associated with Stuttering Persistence: A Meta-Analysis

By

Cara Michelle Singer

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Hearing and Speech Sciences

August 9, 2019

Nashville, Tennessee

Approved:

Robin Jones, Ph.D.

Ellen Kelly, Ph.D.

Katerina Ntourou, Ph.D.

C. Melanie Schuele, Ph.D.

TABLE OF CONTENTS

	Page
LIST OF TABLES	iv
LIST OF FIGURES	v
Chapter	
I. Introduction.....	1
Stuttering chronicity.....	1
Potential clinical characteristics related to stuttering persistence.....	3
Synthesizing the empirical evidence for clinical characteristics of stuttering persistence.....	9
The present study	10
II. Method	12
Eligibility criteria	12
Search strategy	13
Study selection	13
Data management.....	14
Risk of bias assessment.....	15
Effect size synthesis	15
Publication bias assessment	16
III. Results.....	18
Study selection	18
Risk of bias assessment.....	22
Effect size synthesis	24

	Summary of descriptive statistics of clinical characteristics	36
IV.	Discussion	38
	Theoretical connections	40
	Clinical implications	40
	Implications for future research	43
	Caveats	45
V.	Conclusion	46
	REFERENCES	47
	APPENDIX A	58
	APPENDIX B	61

LIST OF TABLES

Table	Page
1. Summary of included studies with related reports listed in parentheses	20
2. Summary descriptive statistics of clinical characteristics for persisting and recovered groups	37
3. Clinical characteristics related to stuttering chronicity (with established and insufficient) evidence organized by descending effect size.....	42

LIST OF FIGURES

Figure	Page
1. Flow chart depicting selection of studies.....	19
2. Risk of bias summary figure. The study confounding source of bias category was only assessed for studies that explored speech-language skills and temperament. Overall risk for a study was rated as moderate if more than two categories were rated as having moderate risk.	23
3. Forest plot comparing risk for stuttering persistence between males and females. As indicated by the diamond falling to the right of the line of no effect, males are at increased risk for persistence compared to females.....	25
4. Contour funnel plot of studies reporting risk related to sex. Asymmetry may be due to publication bias, differences in study methods, and the direction of the relation	26
5. Forest plot comparing mean age at stuttering onset for persistent and recovered children. As indicated by the diamond falling to the right of the line of no effect, stuttering onset occurs later for persistent children than recovered children.....	27
6. Forest plot comparing risk for stuttering persistence between children with and without family history of stuttering. As indicated by the diamond falling to the right of the line of no effect, children with a family history of stuttering are at increased risk for stuttering persistence compared to children with a family history of stuttering. No risk differences were found based on family histories of persistent and recovered stuttering.	29
7. Forest plot comparing mean stuttering behaviors for persistent and recovered children. As evidenced by the positioning of the diamond relative to the line of no effect, persistent children exhibited a higher frequency of stuttering-like disfluencies than recovered children. All other differences were nonsignificant. ...	31
8. Forest plot comparing mean speech-language skills for persistent and recovered children. As indicated by the positioning of the diamonds relative to the line of no effect, persistent children exhibit lower speech-sound accuracy, receptive language and expressive language than recovered children. All other differences were nonsignificant. IPSYN = Index of Productive Syntax (Scarborough, 1990), MLU = Mean Length Utterance, DSS = Developmental Sentence Score (Lee & Canter, 1971), GFTA = Goldman Fristoe Test of Articulation (Goldman & Fristoe, 1999), BBTOP-CI = Bankson-Bernthal Test of Phonology – Consonant Inventory (Bankson & Bernthal, 1990), APP-R = Assessment of Phonological Processes – Revised (Hodson, 1986), PPVT = Peabody Picture Vocabulary Test (Dunn & Dunn,	

1997), TELD = Test of Early Language Development (Hresko, Reid, & Hammill, 1999), Reynell = Reynell Language Development Scale (Reynell, 1983), TACL = Test for Auditory Comprehension of Language (Carrow-Woolfolk, 1999), PLS = Preschool Language Scales (Zimmerman, Steiner, & Pond, 1979), EVT = Expressive Vocabulary Test (Williams, 1997), CDI – MacArthur-Bates Communicative Developmental Inventories (Fenson, Dale, & Reznick, 1993), SPELT = Structured Photographic Expressive Language Test (Dawson, Stout, & Eyer, 2003)..... 34

9. Forest plot comparing temperament for persistent and recovered children. As indicated by the all diamonds crossing the line of no effect, differences are nonsignificant..... 35

CHAPTER I

INTRODUCTION

Approximately 80% of young children who begin stuttering fall below diagnostic thresholds of stuttering (i.e., exhibit recovery) within four to five years of stuttering onset (for an epidemiological review, see Yairi & Ambrose, 2013). To investigate characteristics that may differentiate children who eventually recover from stuttering (recovered) and children who persist in stuttering (persistent), multiple prospective cohort studies have been conducted. Clinical characteristics (e.g., sex, age at onset, family history of stuttering, stuttering severity and speech-language ability) have been explored as potential risk factors of stuttering persistence with varying results (e.g., Ambrose, Yairi, Loucks, Seery, & Throneburg, 2015; Paden, Yairi, & Ambrose, 1999; Spencer & Weber-Fox, 2014). The purpose of this study was to: (1) statistically synthesize data from individual prospective cohort studies to identify clinical characteristics early in development that differentiate recovered and persistent children when both groups are still stuttering, and (2) evaluate methodological characteristics of the studies.

Stuttering Chronicity

Stuttering is a neurodevelopmental communication disorder characterized by stuttering-like disfluencies (e.g., sound-syllable repetitions, whole-word repetitions, and audible or inaudible sound prolongations; Tumanova, Conture, Lambert, & Walden, 2014). Whereas diagnostic criteria for stuttering often vary across studies (for discussion,

see Gordon & Luper, 1992), criteria typically include an elevated frequency of stuttering-like disfluencies (i.e., 2-3% frequency of stuttered disfluencies or above) and/or parent concern for stuttering. Approximately 5-8% of preschool-aged children (e.g., children 2:0-5;11 years of age) exhibit a stuttering disorder, with the onset of childhood stuttering typically occurring when a child is between two and five years of age (Månsson, 2000; Yairi & Ambrose, 2005). Within five years of stuttering onset, approximately 80% of children fall below the diagnostic threshold of stuttering, thereby exhibiting recovery (for epidemiological review, see Yairi & Ambrose, 2013). Yairi and Ambrose (2005) reported that of the children who recovered within five years of onset, 26% recovered within 18 months, 40% recovered within 2 years, and 80% recovered within 3 years.

Although it is just one of many factors to consider when making treatment decisions, the high rate of recovery in childhood stuttering has influenced the decision of *when* and *how* to treat young children who stutter (e.g., Bernstein Ratner, 2018; Nippold, 2018). Onslow and Packman (1999) referred to the interaction between recovery and treatment procedures as “*the* most pressing issue for modern clinicians who have clinical contact with children who begin to stutter” (p. 114). It has long been recommended that speech-language pathologists (SLPs) consider both the length of time the child has been stuttering and the child’s risk for persistence when making treatment decisions for young children (e.g., Kelman & Nicholas, 2008; Zebrowski, 1997). However, assessing a child’s risk for persistence is dependent on the individual SLP’s assessment of prognostic factors of stuttering chronicity within a clinical context.

One source of information that can help SLPs identify such prognostic factors is empirical evidence. Empirical evidence, along with clinical experience and client

preferences, drive evidence-based practice (e.g., Dollaghan, 2007). Empirical evidence on prognostic factors that differentiate persistent and recovered children has come from prognostic studies (e.g., Ambrose et al., 2015; Spencer & Weber-Fox, 2014; Yairi & Ambrose, 2005). These prognostic studies can be characterized as prospective cohort studies in which children who stutter are followed for multiple years to determine which children eventually recover and which children persist. Characteristics of the groups at study entry are then compared to identify potential prognostic factors.

The multifactorial nature of developmental stuttering (for multi-factorial perspectives on stuttering, see Conture & Walden, 2012; Smith & Weber, 2017), and phenomena that differentiate children who stutter from children who do not stutter (e.g., sex differences) have motivated the selection of characteristics that have been explored in longitudinal studies. Overall, important contributors to stuttering include linguistic, speech-motor, temperament, neurological, and genetic factors. Accordingly, these contributors have been explored for their prognostic utility from a variety of different perspectives, including case history information (e.g., Yairi & Ambrose, 2005), physiological measurement (e.g., Zengin-Bolatkale, Conture, Walden, & Jones, 2018), brain morphometric characteristics (Garnett et al., 2018), and behavioral measures (e.g., Ambrose et al., 2015). However, the ability of SLPs to assess and interpret these factors varies (i.e., clinical utility).

Potential Clinical Characteristics Related to Stuttering Persistence

Clinical characteristics are of particular interest in the present study because they have high clinical utility and can be assessed during a comprehensive speech-language

evaluation (for example of an evaluation protocol, see Clark, Choi, & Tumanova, 2017). Clinical characteristics are descriptive features of the child including sex, family history of stuttering, age at stuttering onset, child's demonstrated and described stuttering behaviors, speech-language skills, and temperament. Some of these factors have been assessed for their prognostic value since the 1950's (e.g., Johnson, 1959). These characteristics and their possible contribution to stuttering chronicity will be reviewed individually below. Following this review, potential explanations for the inconsistency in study findings will be discussed in the section "Synthesizing the Empirical Evidence for Clinical Characteristics of Stuttering Persistence".

Sex. Sex differences in children and adults who stutter have been explored. Due to the relatively small sex difference in young children who stutter (Reilly et al., 2009; Shimada, Toyomura, Fujii, & Minami, 2018), but the much larger male-to-female ratio in young adults who stutter (Craig, Hancock, Tran, Craig, & Peters, 2002), males have commonly been thought to be at higher risk for stuttering persistence. For example, Shimada et al. (2018) reported a 1.57:1 male-to-female ratio from a community-based sample of three-year old children. Similarly, Mansson (2005) reported a 1.34:1 male-to-female ratio from a community-based sample of children up to the age of four years. In contrast, Craig et al. (2002) reported a 3.3:1 male-to-female ratio and 4:1 male-to-female ratio in community-based samples of elementary age children who stutter (i.e., six to ten years old) and adolescents and young adults who stutter (i.e., 11 to 20 years old), respectively. However, longitudinal studies have not consistently detected that males are at a greater risk for persistence than females. For example, Ambrose, Cox, and Yairi

(1997) found that males were more likely to persist than females; however, males were not found to be at increased risk for persistence compared to females in other studies (e.g., Johnson, 1959; Kefalianos et al., 2017).

Age at onset. Age at stuttering onset is often based on an estimation of when caregivers first noticed the child beginning to exhibit stuttering in his/her speech. Based on parent estimates, girls have been found to exhibit an earlier onset than boys by approximately five months (Yairi & Ambrose, 1992). The earlier onset exhibited by girls, as well as their potentially higher recovery rate, partially contributes to speculation that a later onset of stuttering increases a child's risk for stuttering persistence. Further, Yairi and Ambrose (2005) speculated that "the older the child, the higher the expectations and various demands on him/her" (p. 349), which may increase the demands on the child's speech-motor system and, thereby, the child's risk for stuttering persistence. Consistent with this speculation, Yairi, Ambrose, Paden and Throneburg (1996) found that persistent children had a later age of onset than recovered children. However, other studies have found no difference in age of onset between persistent and recovered children (e.g., Ambrose et al., 2015; Roehl, 2018).

Family history of stuttering. Jameson (1955) extended the exploration of whether a family history of stuttering may be associated with a child's risk of ever stuttering to whether it may indicate a child's risk of persistence. Since then, a family history of stuttering has continued to be explored as a clinical characteristic of stuttering persistence with some studies finding null results (e.g., Kefalianos et al., 2017; Rommel,

Hage, Kallehne & Johannsen, 2000) and others finding that children with a family history of stuttering are at a greater risk to persist (e.g., Walsh et al., 2018; Yairi & Ambrose, 2005). Over the years, researchers have included more specific measures of family history, including a family history of persistent stuttering versus a family history of recovered stuttering (e.g., Yairi, Ambrose, Paden, & Throneburg, 1996). For example, Yairi et al. (1996) and Walsh et al. (2018) reported that a greater proportion of persistent children had relatives that persisted than recovered children, which indicates that a family history of persistent stuttering may increase a child's risk for stuttering persistence.

Stuttering frequency and severity. Stuttering frequency relates to the rate of occurrence of stuttering-like disfluencies in a child's speech, whereas stuttering severity more globally describes stuttering and often includes evaluations of stuttering frequency, duration, type, and/or associated tension. Master clinicians have suggested that certain aspects of stuttering behaviors such as the presence of tension or prolongations, which often relate to greater stuttering severity, may be associated with an increased risk for stuttering persistence, especially when the child has been stuttering for longer than a year (e.g., Conture, 1990; Zebrowski, 1993). Stuttering frequency is typically assessed by counting the number of stuttering-like disfluencies (SLD; e.g., part- and whole-word repetitions and dysrhythmic phonations) in a speech sample. Stuttering severity can be assessed either by the SLP or the caregivers. Clinician-rated stuttering severity is measured commonly using rating scales (e.g., Yairi & Ambrose, 1999), the Stuttering Severity Instrument (Riley, 1972), or the weighted SLD (Ambrose & Yairi, 1999). Parent-rated stuttering severity is often measured using rating scales (e.g., Curran &

Hood, 1977; Onslow, Andrews, & Costa, 1990; Yairi & Ambrose, 1999). Bostian (2017) found that four- to five-year-old children who went on to persist exhibited more severe stuttering and higher stuttering frequency than similarly-aged children who eventually recovered, but many other studies (that have included children young than four) have reported nonsignificant differences related to stuttering severity and frequency (e.g., Ambrose et al., 2015; Garnett et al., 2018).

Speech-language skills. Speech-language skills are potentially important risk factors for stuttering persistence, and are emphasized in multiple models of stuttering. For example, based on the Dual-Diathesis Stressor Model (Conture & Walden, 2012) relatively low or high speech-language skills increase a child’s vulnerability to stutter and perhaps persist. Similarly, the Multifactorial Dynamic Pathways Theory (Smith & Weber, 2017) proposes that a child’s ability to manage speech-language skills is one factor that impacts the child’s likelihood of recovering or persisting. Although children who persist likely exhibit varying speech-language skills (e.g., some children have high speech-language skills, whereas others have low speech-language skills), there may be “consistent patterns characteristic of the trajectories to persistent stuttering versus recovery from stuttering” (Smith & Weber, 2017; p. 2497). Many studies have reported null results across multiple speech-language skills (e.g., Ryan, 2001; Singer et al., 2019b). However, some studies report that persistent children, compared to recovered children, perform significantly lower on some measures of speech or language, such as speech-sound accuracy (e.g., Spencer & Weber-Fox, 2014) and expressive vocabulary (Ambrose et al., 2015).

Temperament. Temperament also may relate to stuttering chronicity.

Temperament refers to an individual's stable, trait-like characteristics related to their emotional reactivity and regulation (Rothbart, Ahadi, & Hershey, 1994). Children who are generally (i.e., not necessarily in relation to their stuttering) more emotionally reactive and/or less able to regulate their emotions may be at greater risk for stuttering persistence. This possibility would seem to be congruent with theoretical speculation that emotion, for at least some children, may be one important contributor to stuttering (Conture & Walden, 2012). Along with other researchers (e.g., Ambrose et al., 2015), we speculate that temperament may not only contribute to stuttering, but also to whether a child is more likely to persist or recover. Caregiver questionnaires, such as the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001) have been validated to assess temperament in young children. To date, findings have been inconsistent regarding whether negative reactivity (i.e., tendency to react negatively) assessed via caregiver report, is a clinical characteristic related to stuttering persistence. Using the CBQ as a measure of temperament, Ambrose et al. (2015) found that persistent children exhibited higher negative reactivity than recovered children, whereas Zengin-Bolatkale et al. (2018) found no statistically significant differences between the groups using the same measure. Neither study reported differences related to effortful control (e.g., emotional regulation) and surgency¹ (e.g., positive reactivity).

¹ Ambrose et al. (2015) reported nonsignificant between-group findings related to surgency when the persisting, recovered, and non-stuttering groups were compared. When we extracted means and standard deviations of surgency scores for the persisting and recovered groups from Figure 5 in Ambrose et al. (2015), we found that the persistent children exhibited higher surgency scores than the recovered children.

Synthesizing the Empirical Evidence for Clinical Characteristics of Stuttering

Persistence

To assist SLPs in synthesizing findings from prospective cohort studies, reviews (e.g., Clark et al., 2017; Walsh et al., 2018) have been conducted. These reviews have informed evaluation practices related to stuttering persistence. However, a systematic review that includes all available studies has not yet been conducted. For example, a risk factor chart (Guitar & Conture, 2006) was created to help clinicians assess a child's risk for stuttering persistence based on findings from a large cohort study conducted at the University of Illinois. To date, it has not been updated to include subsequent studies. To further facilitate evidence-based practice, an updated synthesis of empirical evidence is warranted based on all available empirical evidence.

A synthesis that employs a meta-analytic approach and includes a critical evaluation of methodological characteristics of included studies would be particularly beneficial. A meta-analytic approach allows for a quantitative synthesis of study findings and provides more power to detect the prognostic value of clinical characteristics compared to any individual cohort study. This issue is particularly important considering that many researchers have cited the small number of persistent and recovered children in the cohort studies as a potential reason for lack of statistical significance (e.g. Spencer & Weber-Fox, 2014; Singer et al., 2019b).

In addition, a critical evaluation of methodological characteristics is also considered “essential to assess and identify biases sufficiently large to distort study results” (Hayden et al, 2013; p. 280). Methodological characteristics (e.g., diagnostic criteria, participant eligibility criteria, length of follow-up, etc.) influence the similarity

between study samples, how representative study samples may be of the target population, and the validity of stuttering outcome classification. Accordingly, there has been speculation that methodological differences between the cohort studies may contribute to the inconsistent findings across the studies (e.g., Clark et al. 2017); however, a critical evaluation has not yet been conducted. By evaluating methodological characteristics and creating eligibility criteria, a review would identify and exclude studies that might interfere with the validity of the results. These characteristics can also be explored as potential moderators of study findings. For example, perhaps differences in the participants' age might explain why some studies detect significant differences in clinical characteristics known to change over time (e.g., sex-ratio, stuttering severity, and stuttering frequency) whereas others do not. Lastly, by evaluating characteristics of studies, it can be determined whether some studies have more inherent sources of bias than others and which sources of bias are most commonly found in prospective cohort studies of young children who stutter.

The Present Study

The present study evaluates the relation of clinical characteristics to stuttering persistence using a meta-analytic approach. This approach can statistically estimate overall differences in clinical characteristics between persistent and recovered children based on all included primary prospective cohort studies, as well as assess whether any methodological characteristics may moderate observed differences between studies. By focusing on clinical characteristics when a child is likely to be early in stuttering development (i.e., younger than six years of age), the present study assessed the

prognostic value of clinical characteristics that can be assessed during a routine, initial evaluation of childhood stuttering. The present study also assessed risk of bias within, and the heterogeneity between, the individual primary prognostic studies. We addressed three research questions:

1. What sources of bias are frequently found in prospective cohort studies of children who stutter that might impact study findings?
2. Are there clinical characteristics that differentiate children who persist in stuttering and children who eventually recover from stuttering in early childhood?
3. Are sex, stuttering severity, and stuttering frequency differences between children who persist in stuttering and children who recover from stuttering moderated by time since onset (TSO) or age at study entry?

CHAPTER II

METHOD

This study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher, Liberati, Tetzlaff, & Altman, 2009) guidelines. The methods were registered online with PROSPERO (CRD42019118590).

Eligibility Criteria

Prospective cohort studies were included in the present study if they met the following criteria. First, studies must have been longitudinal in nature and followed participants for at least 18 months. Second, participants must have been under six years of age at study entry. We selected this criterion to target children near stuttering onset. Third, assessments must have been used to identify participants (a) as stuttering at study entry and (b) as either recovered or persistent at study completion. Fourth, at least one clinical marker must have been assessed at study entry when all participants were stuttering and reported separately for the persistent and recovered groups.

Inclusionary criteria for clinical characteristics were based on clinical utility. Specifically, we included data collected from caregiver report (e.g., interview, questionnaires, scales), normative assessments, children's speech samples, and child report, which are commonly used in assessments of children who stutter (e.g., Clark et al., 2017). Data that require special equipment and are not commonly used by clinicians

in the assessment of childhood stuttering, such as physiological, brain morphometry, and genetic analyses (except for family history and pedigree analysis) were excluded.

Individual reports (e.g., published articles, doctoral dissertations, books, etc.) were included if they reported quantitative results of a potential clinical marker of stuttering persistence that permitted the calculation of an effect size. There were no publication year or language restrictions.

Search Strategy

Studies were identified by several means to minimize publication bias and maximize the likelihood of finding all relevant studies. PsycINFO, ProQuest Dissertations and PubMed were searched in June 2019 using the following search terms: Ti,ab ((stammer* OR stutter*) AND (recover* OR persist* OR longitudinal) AND child*). The first author also searched websites of selected journals (i.e., *Journal of Fluency Disorders*, *Journal of Speech, Language and Hearing Research*, *Journal of Communication Disorders*, *American Journal of Speech-Language Pathology*, *International Journal of Speech-Language Pathology*, *International Journal of Language and Communication Disorders*) and the American Speech-Language Hearing Association using key terms. Authors were contacted if additional information was needed.

Study Selection

Research reports from these search methods were narrowed down during a two-step double-screened review. The first author and a reliability coder reviewed the title and abstract of these reports using Microsoft Excel. Any report that compared children who

stutter and persisted (i.e., Persistent group) and children who stutter and recovered (i.e., Recovered group) was included in the next stage of the review process. In this second stage, the first author and a reviewer read each report to identify whether it was eligible or to identify reason(s) for ineligibility. During both steps, disagreements were discussed until a consensus was reached. Lastly, the reference lists of included reports were searched to identify additional eligible reports.

Data Management

For all reports that met inclusionary criteria, the first author and a reviewer extracted and entered all data in a Microsoft Excel spreadsheet. Data extracted from reports included: author, year of publication, funding source, journal, language, document type, recruitment source, eligibility requirements, time-since-onset criteria at study entry, group classification criteria, length of follow-up, percent participants treated, criteria for length of recovery, sample sizes, mean ages, sex distributions of the persistent and recovered groups, reporting of attrition, and method to control for confounding variables. The coders then recorded data for each clinical characteristic reported in each study. When data were reported for multiple visits, only data obtained closest to stuttering onset were extracted². When multiple measures of a characteristic were reported (e.g., clinician-reported stuttering severity and parent-report stuttering severity) all data were extracted.

² The only exception to this rule were data extracted from Yairi and Ambrose (2005). Data collected less than 12 months post onset were extracted instead of data collected 0-6 months post onset because fewer children were observed at the earlier timepoint.

Risk of Bias Assessment

To explore the first research question, risk of bias stemming from study characteristics was assessed with the Quality in Prognostic Studies tool (QUIPS; Hayden, van der Windt, Cartwright, Côté, & Bombardier, 2013). For each study, the first author and a reviewer used the QUIPS to individually assign ratings of low, moderate or high risk of bias to each of six domains: study participants, prognostic factor measurement, study attrition, outcome measurement, study confounding, and statistical analysis and reporting was evaluated. For any disagreement, the coders discussed the domain until a consensus was reached. An a priori decision was made that studies rated as being at high risk for bias would be excluded.

Further, characteristics related to the risk of bias were assessed for each report. When two reports with participant overlap reported similar data, data from the report with the lowest level of bias across the six domains were extracted for analysis. For example, there was considerable participant overlap in Roehl (2018) and Spencer and Weber-Fox (2014). Data from Roehl (2018) were extracted for receptive vocabulary over data from Spencer and Weber-Fox (2014) due to group classification being based on longer length of follow-up (i.e., lower risk of bias related to outcome measurement).

Effect Size Synthesis

Data were analyzed using the *metafor* package in R. Effect sizes were calculated for each individual characteristic (e.g., speech-sound skills, receptive language) that was measured in at least two studies. Separate measures of an individual characteristic (e.g., clinician-reported stuttering severity and parent-reported stuttering severity) were

analyzed independently of one another whenever possible. Thus, multiple effect sizes were calculated sometimes from a single study.

To explore our second research question, overall effect sizes were synthesized using a random effects model. A random effects model assumes there is between- and within-study heterogeneity that results in multiple true population effect sizes. Effect sizes from individual studies were weighted using inverse variance weights that are sample-size dependent; studies with greater sample sizes receive more weight. A risk ratio was calculated to estimate differences related to sex and the presence/absence of a known family history of stuttering. Hedges' g (Hedges, 1982), a standardized mean difference that corrects for small sample size, was calculated to estimate differences related to continuous measures (e.g., speech-language scores).

Homogeneity tests were conducted using the Q , I^2 , and τ^2 statistics. The Q statistic reflects the amount of heterogeneity; I^2 and τ^2 reflect the *proportion* and the *amount* of true heterogeneity, respectively. I^2 values of 25%, 50%, and 75% are considered as low, moderate and high proportions of heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003).

Publication Bias Assessment

Lastly, publication bias was evaluated to determine whether findings may be influenced by including only studies that reported larger than average effects, which are more likely to be published. When at least ten studies were included in a meta-analysis on a given clinical characteristic, publication bias was investigated by creating funnel plots

and conducting an Egger regression test for funnel plot asymmetry (Egger, Smith, Schneider, & Minder, 1997).

CHAPTER III

RESULTS

Study Selection

Study selection results are detailed in Figure 1. From an initial pool of 382 reports, 88 reports passed the title and abstract review. The interrater reliability index, prior to reaching consensus, was .93 for this stage of the process. Exclusionary criteria were evaluated in the order listed in Figure 1. The full text review resulted in 39 reports meeting eligibility criteria; interrater reliability index, prior to reaching consensus, was .94 for the full-text stage. Ultimately, these reports described 11 unique studies (i.e., non-overlapping participant samples) that met eligibility requirements for the present study. Descriptions of these studies can be found in Table 1.

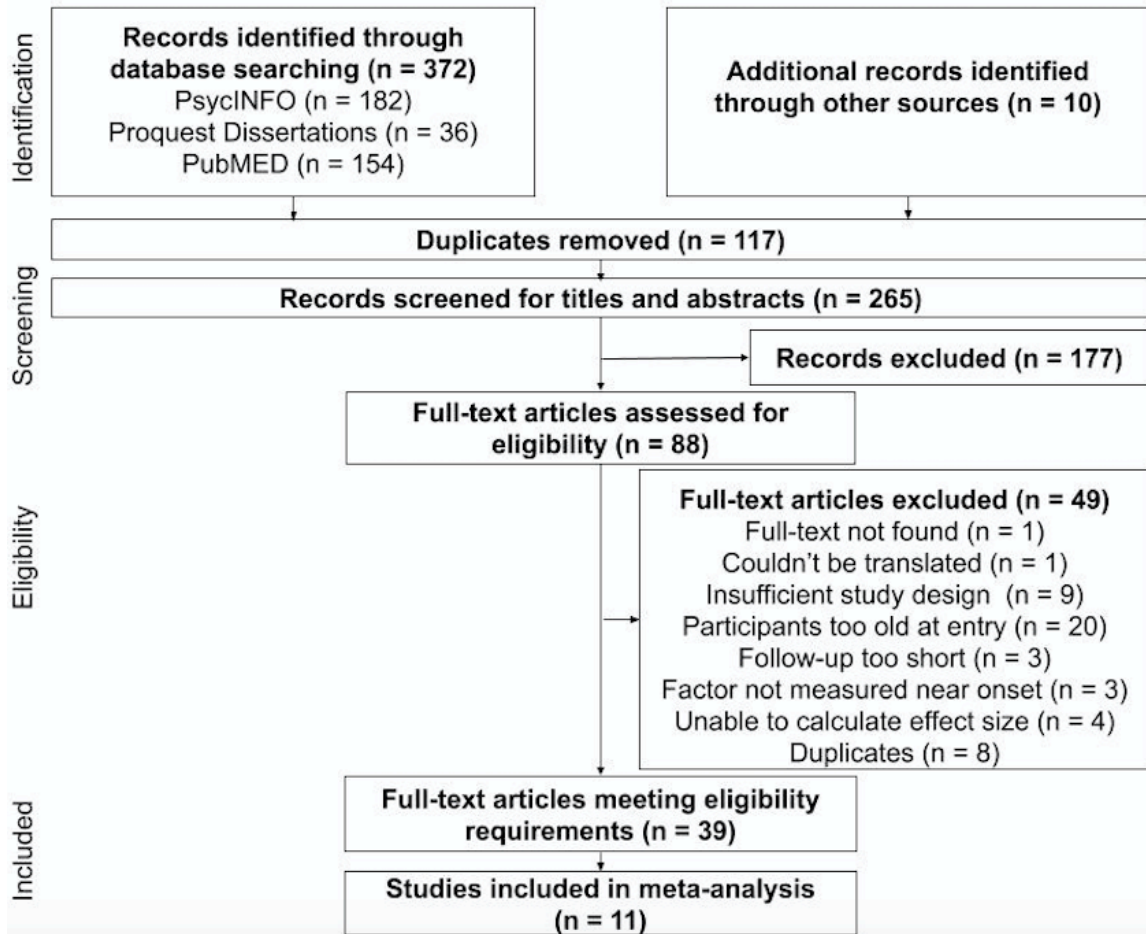


Figure 1. Flow chart depicting selection of studies.

Table 1

Summary of included studies with related reports listed in parentheses

Included Studies		Stuttering-Related Measures				Attrition	Participants		
Study (Reports)	Country	Entry age (yr.)	Entry TSO (mo.)	Criteria for Group Classification	Follow-Up Length		Recruitment	Inclusionary/ Exclusionary Criteria	Percent Receiving Treatment
Outcomes of Children with Hearing Loss (Arenas, Walker, & Oleson, 2017)	USA	3	Freely vary	Parent report	at least 2 years or until 8 years old	R	Multi-site	Hearing loss between 25 and 75 dB	34.8% of the recovered group and 40% of the persistent group
Twins Early Development Study (Dworzynski, Remington, Rijdsdijk, Howell, & Plomin, 2007)	England	≤ 5	< 12	Parent report	until 7 years old	R	Population-based	Twins	NR
Erasmus University Medical Centre Study (Franken, Koenraads, Holtmaat, & van der Schroeff, 2018)	Netherlands	3-5	Freely vary	3% SS; SSI > 9; parent and clinician-rated stuttering severity of at least 2 on eight-point scale	up to 9 years	R	Clinical	No known neurological, intellectual or psychosomatic problems	91% of recovered group and 100% of persistent group
Illinois Stuttering Research Project (Ambrose et al., 1997; Hall, Amir, & Yairi, 1999; Mahurin-Smith & Ambrose, 2013; Paden, Ambrose, & Yairi, 2002; Paden & Yairi, 1996; Paden et al., 1999; Sawyer, 2005; Subramanian, Yairi, & Amir, 2003; Throneburg & Yairi, 2001; Watkins & Yairi, 1997; Watkins, Yairi, & Ambrose, 1999; Yairi & Ambrose, 1992, 1999, 2005; Yairi et al., 1996)	USA	freely vary	< 12	3% SS; parent- and clinician-rated stuttering severity of at least 2 on eight-point scale	4 years	R	Single-site	No known neurological disorders	0% of recovered group and 89% of persistent group (Yairi & Ambrose, 2005)
Early Language in Victoria Study (Kefalianos et al., 2017)	Australia	2	< 12	Stuttering severity rating of at least 2 on a 10-point scale by parents and clinician; 12 months for recovery	at least 3 years	R	Population based	NR	13.4% of recovered group and 16.7% of persistent group

Netherlands – High Risk (Kloth, Kraaimaat, Janssen, & Brutten, 1999)	Netherlands	2-5	< 12	parent report	at least 4 years	R	Single-site	Positive parental history of stuttering	NR
Purdue Stuttering Project (Bostian, 2017; Hilger et al., 2017; Kreidler et al., 2017; Leech et al., 2017; Roehl, 2018; Spencer & Weber-Fox, 2014; Walsh et al., 2018)	USA		Freely vary	3% SS; clinician-rated stuttering of severity 2 or higher on eight-point scale	2-5 years	NR	Multi-site	No known neurological disorders or injury	Only treatment prior to study reported
Ulm Study (Brosch, Hage, & Johannsen, 2001; Brosch, Hage, Kalehne, & Johannsen, 1999; Hage, 2001; Rommel, Hage, Kalehne, & Johannsen, 2001)	Germany	3-6	Freely vary	NR	3 years	R	Clinical	NR	Treatment allowed; specifics not reported
California- Long Beach (Ryan, 2001)	USA	< 6	Freely vary	3.0 SW/min; trends of SW/min	up to 10 years	NR	Clinical	No previous treatment	Treatment reported for persistent children; not recovered
Subtypes and Risk Factors in Childhood Stuttering Study (Ambrose et al., 2015; Buhr, 2007; Hollister, Van Horne, & Zebrowski, 2017)	USA	freely vary	< 12	3% SS; parent and clinician stuttering severity; recovery needed to be exhibited for 12 months based on parent report	up to 5 years	NR	Multi-site	No known neurological disorders	Only reported for small subset (e.g., Hollister et al., 2017)
Vanderbilt's Developmental Stuttering Project (Erdemir, Walden, Jefferson, Choi, & Jones, 2018; Singer, Walden, & Jones, 2019a, 2019b; Zengin-Bolat kale et al., 2018)	USA	3-5	Freely vary	3% SW and above 10 on SSI; recovery required absence of parent concern	at least 2 years	R	Single site	No known neurological conditions and scores above 17 th percentile on speech-language measures	23% of recovered group and 20% of persistent group

Note: R = Reported; NR = Not Reported; SS = stuttered syllables; SW = stuttered words; SSI = Stuttering Severity Instrument (Riley, 1972).

Risk of Bias Assessments

Results pertaining to the risk of bias assessment for each study can be found in Figure 2. The most common sources of bias identified across studies included study participation that limited generalizability of findings based on inclusionary/exclusionary criteria, limited follow-up (i.e., less than three years) which could impact group classification, insufficient reports of attrition, and lack of reports on methods to control for the effect of known confounding variables. Four of the studies (36%) had eligibility criteria that negatively impacted how representative the sample was of the general population of children who stutter. Specifically, Arenas et al. (2017) only included participants with hearing loss, Kloth et al. (1999) only included children who had at least one parent with a history of stuttering, Dworzynski et al. (2007) only included twins, and the Vanderbilt Developmental Stuttering Project (Erdemir et al., 2018; Singer et al., 2019a; 2019b; Zengin-Bolatkale et al., 2018) only included children who scored above 17th percentile on assessments of speech and language skills (excluding stuttering). Of the seven studies that reported data on attrition, none reported between-group analyses of differences between children who did and did not complete the study. Last, of the nine studies that explored the prognostic value of speech-language skills and/or temperament, six studies did not consistently report methods used to control for known confounding factors. None of the studies were determined to exhibit high risk of bias. So all eligible data were included in subsequent analyses.

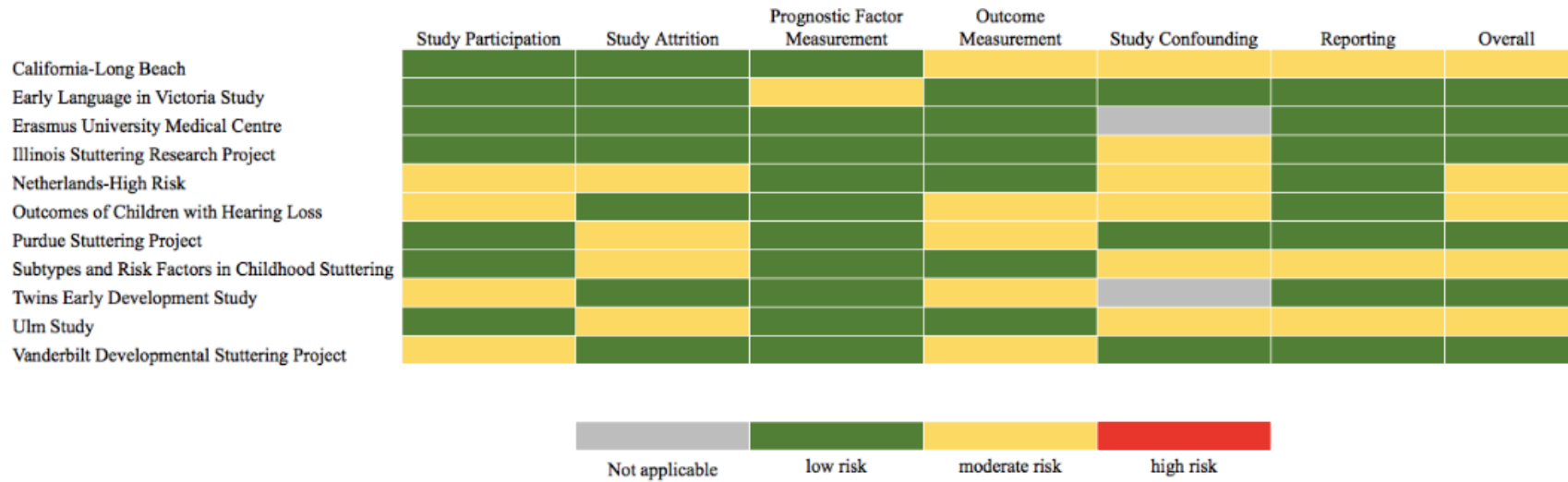


Figure 2. Risk of bias summary figure. The study confounding source of bias category was only assessed for studies that explored speech-language skills and temperament. Overall risk for a study was rated as moderate if more than two categories were rated as having moderate risk.

Effect Size Synthesis

Based on the nature, availability, and quality of the clinical characteristics included in the 11 studies, study outcomes were grouped into the following categories: sex, age at onset, family history of stuttering, stuttering behaviors, speech-language behaviors, and temperament. The results described below are divided and reported by category. Forest plots are included for each category. The location of the box along the x-axis indicates an individual study's effect size, the length of the lines connected to the box indicate the 95% confidence interval (CI), and the size of the box represents the individual study's weight in the analysis. The location of the center of the diamond along the x-axis indicates the estimated mean effect size of all the studies when combined, and the width of the diamond represents the 95 CI. If the diamond crosses the dashed line (i.e., the line of no effect), no effect is indicated. Heterogeneity statistics are reported within the forest plots. Detailed information for each report can be found in Appendix A. Moderator and publication bias analyses are reported if there was a sufficient number of studies to conduct such analyses.

Sex. As shown in Figure 3, males are 1.48 times more likely to persist than females (*risk ratio* = 1.48, 95 CI [1.10, 2.00]). Low heterogeneity was detected across the eleven studies ($I^2 = 22.4\%$, $\tau^2 = 0.05$); therefore, follow-up meta-regressions were not conducted.

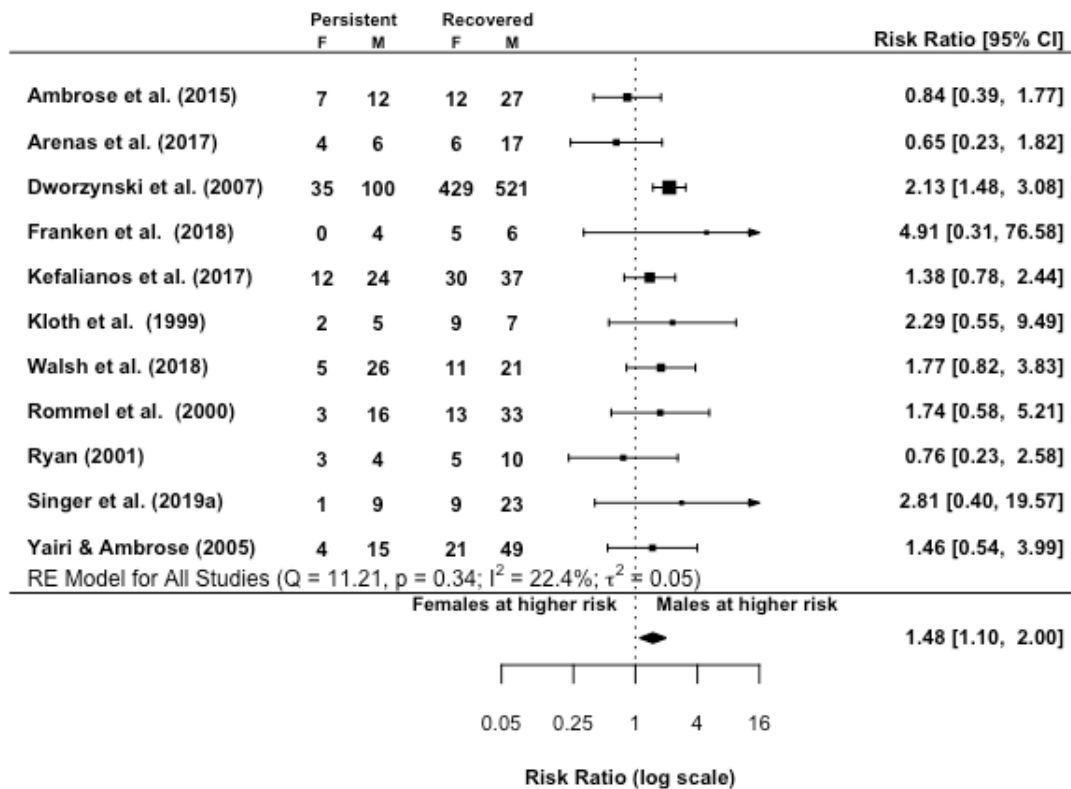


Figure 3. Forest plot comparing risk for stuttering persistence between males and females. As indicated by the diamond falling to the right of the line of no effect, males are at increased risk for persistence compared to females.

As shown in Figure 4, a funnel plot was constructed to explore possible publication bias. There was no evidence of small study effects or publication bias based on an Egger's regression test ($z = -0.035, p = 0.972$).

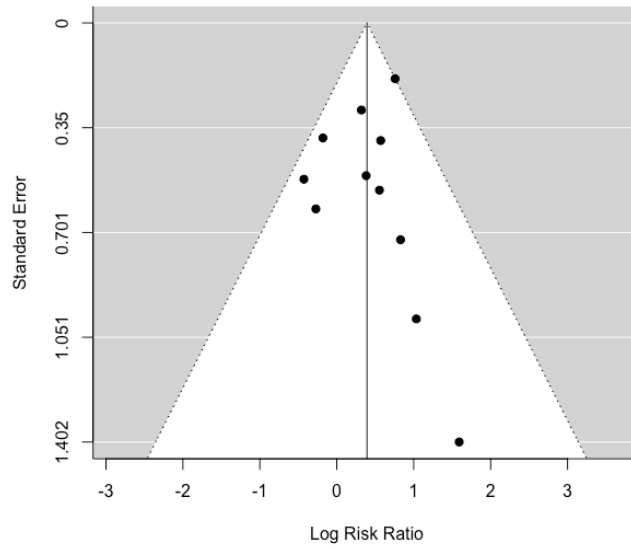


Figure 4. Contour funnel plot of studies reporting risk related to sex. Asymmetry may be due to publication bias, differences in study methods, and the direction of the relation.

Age at Onset. As seen in Figure 5, persistent children are reported to have a later onset than recovered children, (Hedges' $g = 0.43$, 95 CI [0.16, 0.71]). A low level of heterogeneity was observed across the six studies ($I^2 = 0.0\%$, $\tau^2 = 0.00$).

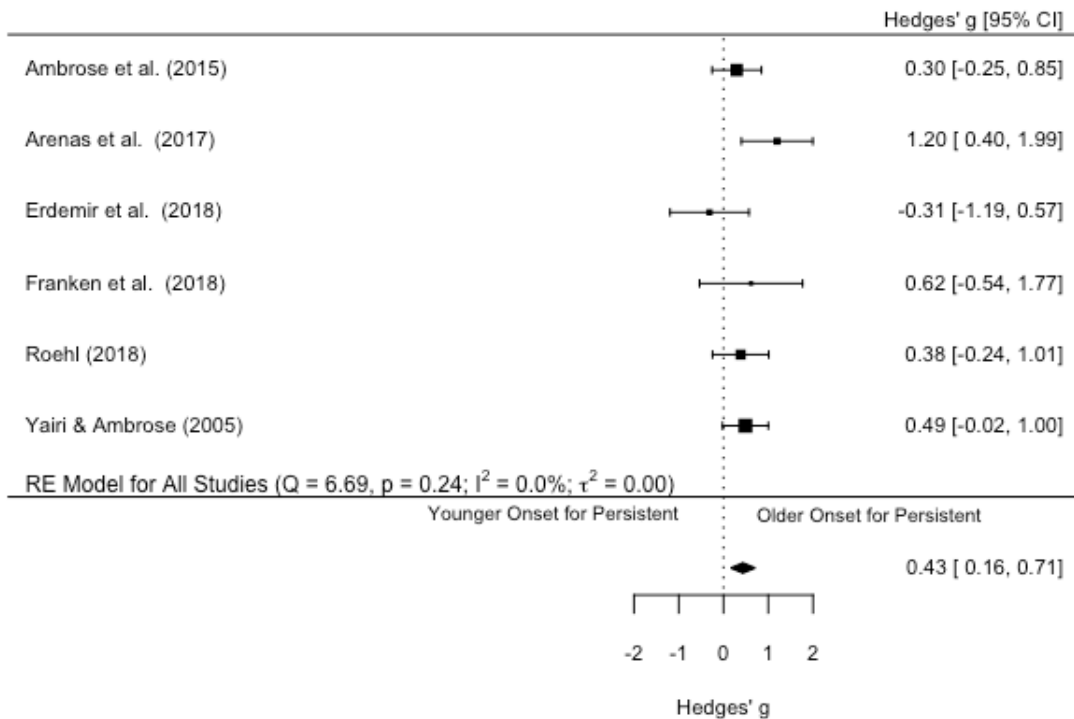


Figure 5. Forest plot comparing mean age at stuttering onset for persistent and recovered children. As indicated by the diamond falling to the right of the line of no effect, stuttering onset occurs later for persistent children than recovered children.

Family History of Stuttering. Three types of family history of stuttering were analyzed separately: any stuttering (persistent and/or recovered), persistent stuttering, and recovered stuttering. The nature of data collection on family history of stuttering varied across the studies: Yairi and Ambrose (2005) and Kefalianos et al., (2017) collected data on immediate family members, Walsh et al. (2018) collected data on first- and second-degree relatives, Franken et al. (2018) collected data on any known family members, and Rommel et al. (2000) did not report specifics.

For all analyses, data reported on the proportion of the children in the persistent and recovered groups with and without family history were included. Data from Yairi and Ambrose (2005) specific to family history of persistent and recovered stuttering pertained

to the proportion of *relatives* that exhibited either persistent or recovered stuttering and therefore was excluded. In contrast, data related to family history of any stuttering reported in Yairi and Ambrose (2005) were based on the proportion of children in each group with a family history of stuttering and were included.

As seen in Figure 6, children with a family history of any stuttering are 1.89 times more likely to persist than children without a family history of stuttering, (*risk ratio* = 1.89, 95 CI [1.27, 2.82]). Low heterogeneity was detected across this model ($I^2 = 0.0\%$, $\tau^2 = 0.00$). There were not significant differences detected in risk for persistence between children with and without a family history of persistent or recovered stuttering, but high heterogeneity was detected across both models, (I^2 range: 45.0 – 51.2%; τ^2 range: 1.06 – 1.10). Due the small number of studies, meta-regressions could not be conducted.

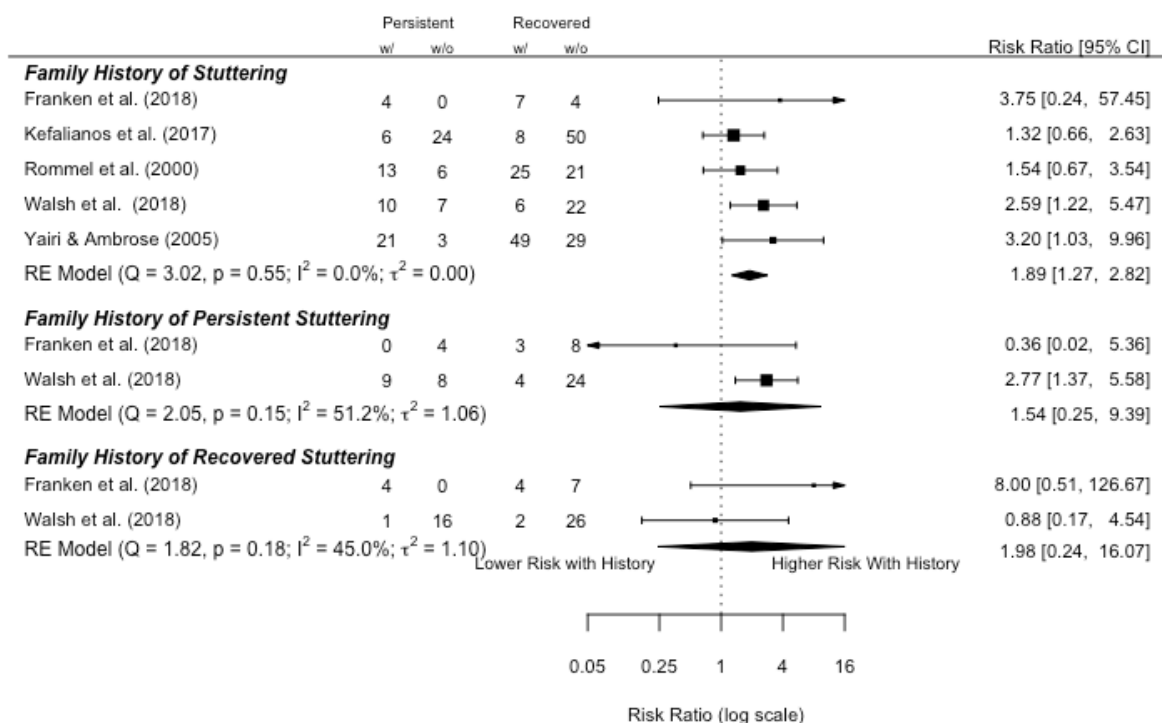


Figure 6. Forest plot comparing risk for stuttering persistence between children with and without family histories of stuttering. As indicated by the diamond falling to the right of the line of no effect, children with a family history of stuttering are at increased risk for stuttering persistence compared to children without a family history of stuttering. No risk differences were found based on family history of persistent and recovered stuttering.

Stuttering Behaviors. Seven stuttering behaviors were analyzed separately: clinician-rated stuttering severity, parent-rated stuttering severity, and frequency of stuttering-like disfluencies, non-stuttering-like disfluencies, part-word repetition disfluencies, single-syllable whole-word repetition disfluencies, and dysrhythmic phonations (i.e., prolongations). To assess a child’s stuttering frequency, studies utilized various methods to collect child speech samples, as described in Appendix B.

As seen in Figure 7, persistent children produced a higher frequency of stuttering-like disfluencies than recovered children (Hedges’ $g = 0.43$, 95 CI [0.18, 0.87]). Low

heterogeneity was detected across the five studies ($I^2 = 27.2\%$, $\tau^2 = 0.04$). Persistent children did not differ in stuttering severity based on clinician³ or caregiver assessment, the frequency of non-stuttering-like disfluencies, part-word repetition disfluencies, single-syllable whole-word repetition disfluencies, or dysrhythmic phonations compared to recovered children. Low heterogeneity was detected across studies reporting data for stuttering severity and nonstuttering-like disfluencies ($I^2 = 0.00$, $\tau^2 = 0.00$). Moderate-to-high heterogeneity was detected across studies reporting data for individual stuttering-like disfluency types (I^2 range: 63.5 – 77.0%; τ^2 range: 0.15 – 0.31), but due to the small number of studies, moderator analyses were not conducted.

³ Two studies (Franken et al., 2018; Roehl, 2018) reported data on two measures of clinician-rated stuttering severity. Due to the small number of studies included in the analysis, robust variance estimation procedures (Tipton, 2015) could not be used and only data on one measure per study could be included (i.e., the first measure reported).

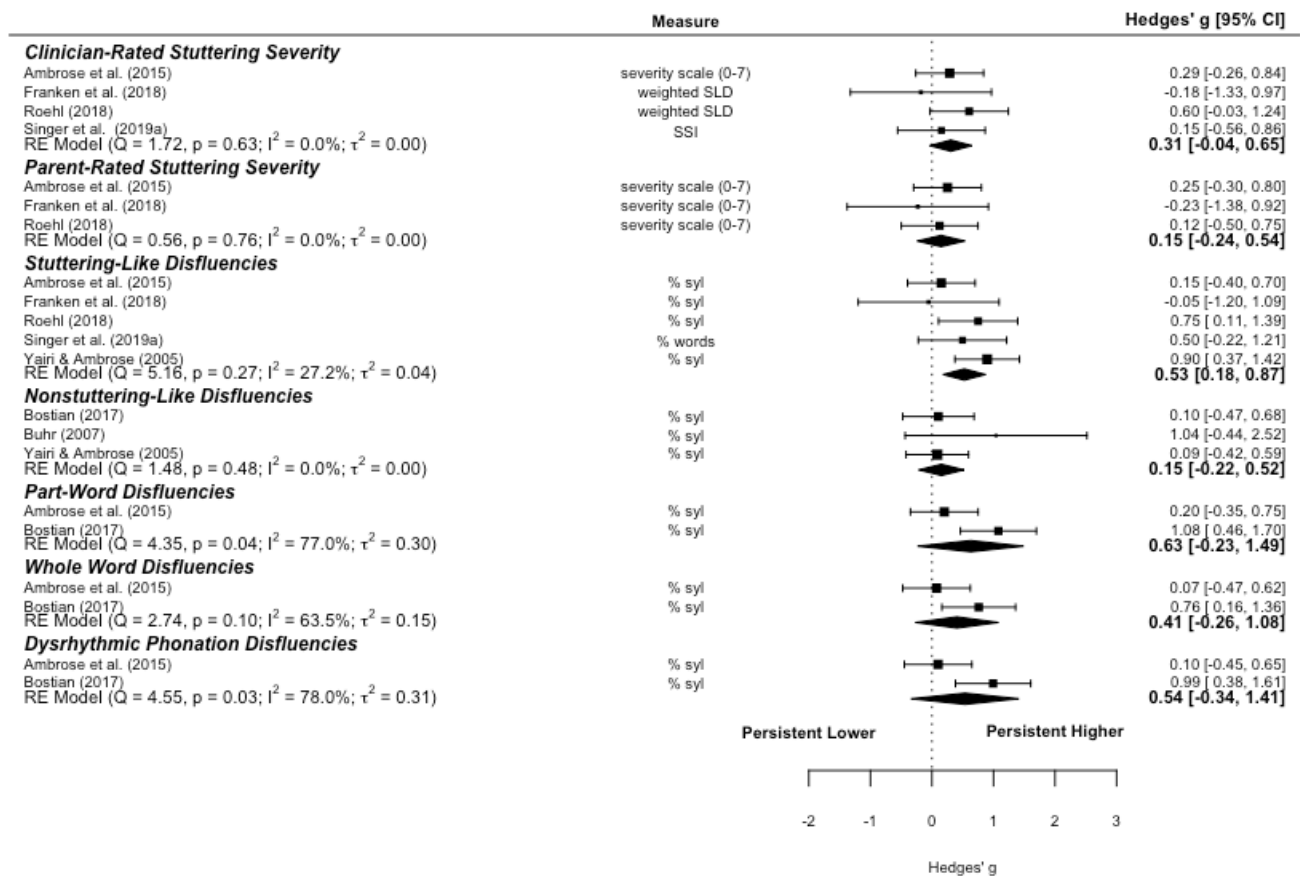


Figure 7. Forest plot comparing mean stuttering behaviors for persistent and recovered children. As evidenced by the positioning of the diamond relative to the line of no effect, persistent children exhibited a higher frequency of stuttering-like disfluencies than recovered children. All other differences were nonsignificant.

Speech-Language Behaviors. Five speech-language skills were quantified via norm-referenced measures of speech-language: speech-sound accuracy, receptive vocabulary, receptive language, expressive vocabulary, and expressive language. Three speech-language skills were quantified with language sample analysis: mean length of utterance (MLU), Index of Productive Syntax (IPSYN; Scarborough, 1990), and Developmental Sentence Scoring (DSS; Lee & Canter, 1971).

As seen in Figure 8, persistent children scored significantly lower on measures of speech-sound accuracy, (Hedges' $g = -0.54$, 95 CI [-0.88, -0.21]), receptive language (Hedges' $g = -0.46$, 95 CI [-0.77, -0.17]), and expressive language (Hedges' $g = -0.43$, 95 CI [-0.69, -0.16]) than recovered children. Low heterogeneity was detected across these models (I^2 range = 0.0 – 10.4%, τ^2 range = 0.00 – 0.01). Persistent and recovered children did not differ significantly on measures of receptive vocabulary or expressive vocabulary. Moderate heterogeneity was detected across these models (I^2 range = 59.5 – 61.9%, τ^2 range = 0.12 – 0.21), but due to the small number of studies, moderator analyses were not conducted.

Sensitivity analyses were conducted to explore whether findings change when data evaluated to be at greater risk for bias related to speech-language behaviors are excluded. Data from Singer et al. (2019a) was evaluated to have moderate risk of bias due to excluding children who scored below the 17th percentile on any one speech-language measure. Similarly, data from Kefalianos et al. (2017) was also evaluated to have moderate risk of bias due to the speech-language evaluation occurring when participants were two years of age, which might have been before some children began stuttering. Results from the sensitivity analyses indicated that significant findings did not change. Persisting children scored lower on measures of speech-sound accuracy (Hedges' $g = -0.62$, 95 CI [-0.99, -0.24], receptive language (Hedges' $g = -0.43$, 95 CI [-0.77, -0.08]), and expressive language (Hedges' $g = -0.44$, 95 CI [-0.73, -0.15]) than recovered children. Further, persisting children did not differ significantly from recovered children on measures of receptive vocabulary (Hedges' $g = -0.31$, 95 CI [-0.82, 0.20]). When

expressive vocabulary data from Singer et al. (2019a) and Kefalianos et al. (2017) were excluded, there was an insufficient number of studies to conduct a meta-analysis.

The length of language samples analyzed varied across studies: Kloth et al. (1999) analyzed 10-min speech samples, Hollister et al. (2017) analyzed 100-utterance speech samples, Buhr (2007) analyzed 50-utterance speech samples, and Watkins et al. (1999) analyzed 250- to 300-utterance speech samples. Further, Hollister et al. (2017) hand calculated IPSYN, whereas Kloth et al. (1999) used a computer program.

As seen in Figure 8, persistent and recovered children, on average, did not exhibit significantly different expressive language skills when quantified using IPSYN, MLU, or DSS. Low heterogeneity was found across these models ($I^2 = 0.00\%$, $\tau^2 = 0.00$).

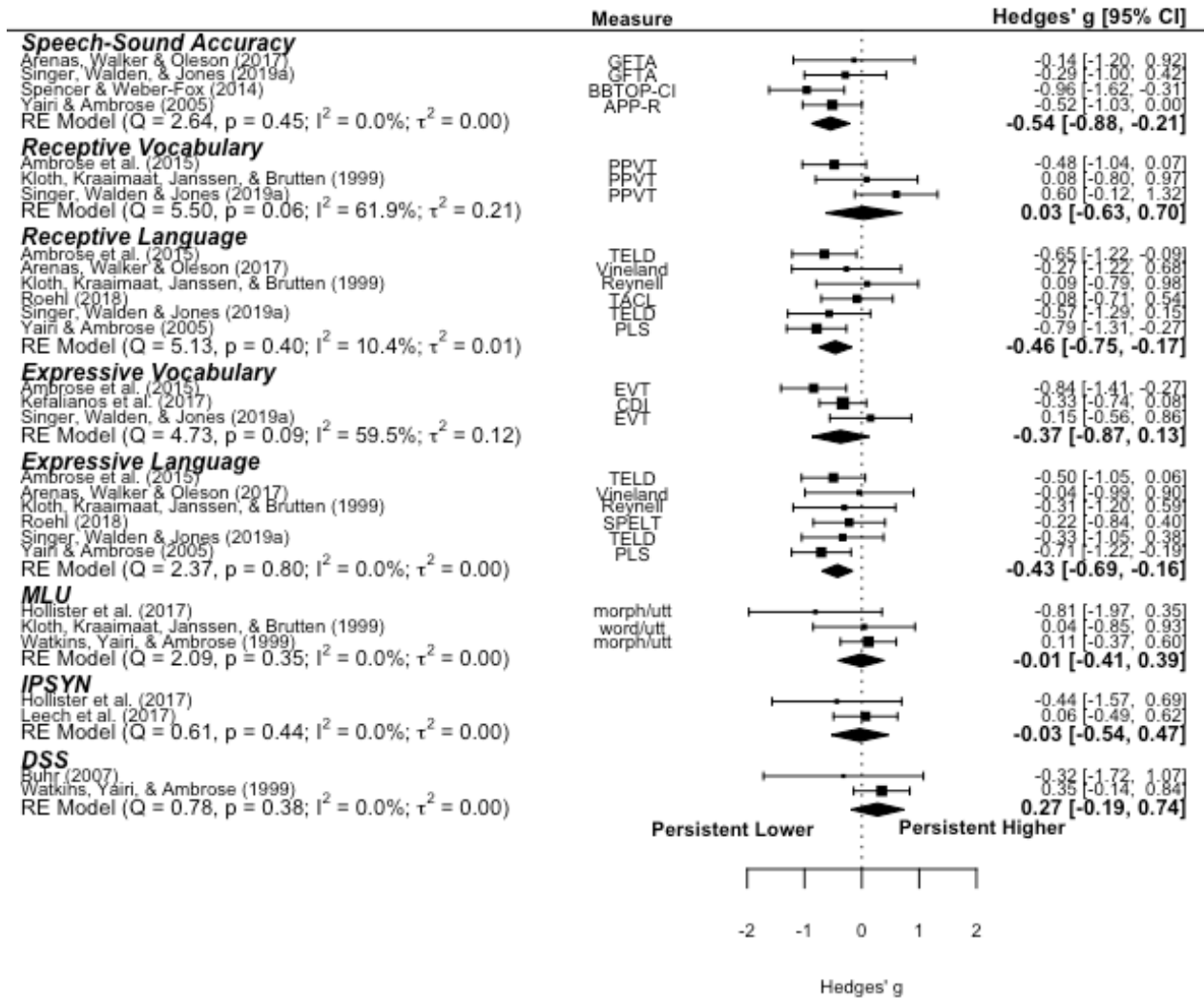


Figure 8. Forest plot comparing mean speech-language skills for persistent and recovered children. As indicated by the positioning of the diamonds relative to the line of no effect, persistent children exhibit lower speech-sound accuracy, receptive language and expressive language than recovered children. All other differences were nonsignificant. IPSYN = Index of Productive Syntax (Scarborough, 1990), MLU = Mean Length Utterance, DSS = Developmental Sentence Score (Lee & Canter, 1971), GFTA = Goldman Fristoe Test of Articulation (Goldman & Fristoe, 1999), BBTOP-CI = Bankson-Bernthal Test of Phonology – Consonant Inventory (Bankson & Bernthal, 1990), APP-R = Assessment of Phonological Processes – Revised (Hodson, 1986), PPVT = Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), TELD = Test of Early Language Development (Hresko, Reid, & Hammill, 1999), Reynell = Reynell Language Development Scale (Reynell, 1983), TACL = Test for Auditory Comprehension of Language (Carrow-Woolfolk, 1999), PLS = Preschool Language Scales (Zimmerman, Steiner, & Pond, 1979), EVT = Expressive Vocabulary Test (Williams, 1997), CDI – MacArthur-Bates Communicative Developmental Inventories (Fenson, Dale, & Reznick, 1993), SPELT = Structured Photographic Expressive Language Test (Dawson, Stout, & Eyer, 2003).

Temperament. Three temperament characteristics were explored: negative affectivity, surgency, and effortful control. These characteristics were reported in Ambrose et al. (2015) and Zengin-Bolat kale et al. (2018) using the CBQ (Rothbart et al., 2001). Temperament data from Kefalianos et al. (2017), measured using the approach/withdrawal scale of the Short Temperament Scale for Children (Sanson, Smart, Prior, Oberklaid, & Pellow, 1994), were determined to be too dissimilar from the other data to be included in an analysis.

As seen in Figure 9, persistent and recovered children did not significantly differ in negative affectivity, surgency, and effortful control. High heterogeneity was detected across the studies for negative affectivity and surgency (I^2 range: 95.2 – 98.5%; τ^2 range: 2.98 – 16.67), but moderator analyses could not be conducted due to the small number of studies. Low heterogeneity was detected across the studies for effortful control ($I^2 = 0.00$, $\tau^2 = 0.00$).

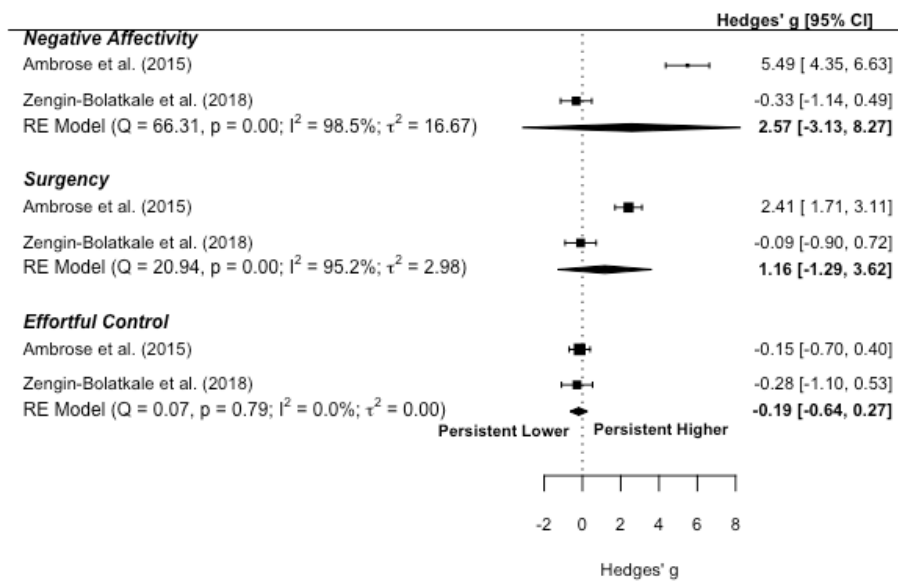


Figure 9. Forest plot comparing temperament for persistent and recovered children. As indicated by all diamonds crossing the line of no effect, differences are nonsignificant.

Summary of Descriptive Statistics of Clinical Characteristics

Refer to Table 2 for the descriptive statistics of clinical characteristics for the persistent and recovered groups. Only data that could be combined (e.g., standard scores could not be combined with raw scores) were used to calculate these descriptive statistics, and, therefore, represent a subset of the data included in the meta-analyses.

Table 2

Summary Descriptive Statistics of Clinical Characteristics for the Persistent and Recovered Groups

Clinical Characteristics	Persistent Group		Recovered Group	
	M (SD)	N	M (SD)	N
<i>Sex (% male)</i>	74.4	297	57.7	1301
<i>Age at Onset (months)</i>	39.6 (13.1)	78	34.4 (8.2)	180
<i>Family History of Stuttering</i>				
Any Stuttering (% positive)	69.2	78	46.8	237
Persisting Stuttering (% positive)	42.8	21	18	39
Recovered Stuttering (% positive)	23.8	21	15.4	39
<i>Speech-Language Behaviors</i>				
Speech-Sound ^{†‡} (standard score)	90.0 (17.0)	34	100.5 (16.1)	64
Receptive Language ^{†‡} (standard score)	111.8 (15.9)	64	123.1 (17.8)	167
Expressive Language ^{†‡} (standard score)	105.5 (15.4)	64	115.9 (18.8)	167
Receptive Vocabulary [‡] (standard score)	107.6 (17.7)	29	109.7 (11.6)	71
Expressive Vocabulary [§] (standard score)	106.9 (14.0)	29	114.1 (11.8)	71
DSS	6.6 (1.5)	26	6.1 (1.5)	66
IPSYN	78.1 (10.6)	36	79.5 (9.2)	27
MLU [‡] (morphemes/utterance)	4.0 (0.9)	30	4.0 (1.0)	67
<i>Stuttering Behaviors</i>				
Stuttering Frequency [†] (% syllable)	8.7 (7.7)	62	6.1 (4.8)	157
Non-Stuttering Frequency (% syllable)	4.8 (2.0)	42	4.8 (2.3)	103
Part-Word Repetitions (% syllable)	3.6 (3.4)	38	2.6 (3.0)	68
Whole-Word Repetitions (% syllable)	3.0 (3.0)	38	2.5 (1.8)	68
Dysrhythmic Phonations (% syllable)	2.4 (2.4)	38	1.6 (1.8)	68
Weighted SLD	11.9 (7.1)	20	9.0 (9.1)	37
Parent Rating Scale (0-7)	4.0 (7.5)	39	3.5 (1.1)	76
<i>Temperament</i>				
Effortful Control	5.0 (0.5)	27	5.1 (0.4)	61
Negative Affectivity	4.2 (0.3)	27	3.8 (0.3)	61
Surgency	4.8 (0.4)	27	4.7 (0.5)	61

Note. Weighted means and standard deviations reported. DSS = Developmental Sentence Score (Lee & Canter, 1971); IPSYN = Index of Productive Syntax (IPSYN; Scarborough, 1990); MLU = Mean Length Utterance; Weighted SLD = Weighed Stuttering-Like Disfluent (Ambrose & Yairi, 1999).

[†] Denotes Arenas et al. (2017) raw data excluded. [‡] Denotes Kloth et al. (1999) non-standardized data excluded. [§] Denotes Kefalianos et al. (2017) raw data excluded. [†] Denotes Singer et al. (2019a) data based on percent words excluded.

CHAPTER IV

DISCUSSION

The present study was the first quantitative synthesis to identify clinical characteristics that differentiate children up to six years of age who recover from stuttering from children who persist in early childhood. Based on the small number of studies available for each individual model, findings are preliminary in nature, but nevertheless represent the *highest level of evidence* for clinical characteristics related to stuttering persistence to date. Children who persist, compared to children who eventually recover, were found: 1) more likely to be male, 2) to begin stuttering at a later age, 3) to have a known family history of stuttering (persistent and/or recovered), 4) to produce a higher stuttering frequency, and 5) to perform lower on measures of speech-sound accuracy, expressive language, and receptive language. The modest effect sizes of these between-group differences and risk ratios can be interpreted to suggest that persistent children, as a group, exhibit vulnerabilities and/or characteristics that might confer heightened risk for the development of chronic stuttering.

It is possible that additional differences will be detected if and when more data become available. For example, findings related to negative and positive emotionality, individual stuttering-like disfluencies types, expressive vocabulary, and family history of persistent stuttering were based on only two or three studies—with one study reporting a significant difference in each case. Further, given the high heterogeneity within these

models, the evidence to support the association between these characteristics and stuttering persistence may be characterized as *insufficient*.

Overall, findings confirm, but also extend previous non-systematic reviews (e.g., Clark et al, 2017; Guitar & Conture, 2006) of potential risk factors for stuttering persistence by providing novel evidence and insight. First, whereas the characteristics (e.g., age at onset, male sex) identified to be associated with stuttering persistence have been reported in previous reviews, the present study provides clear, unequivocal *empirical evidence* supporting the associations between these characteristics and stuttering persistence unlike prior reviews. In other words, the present study validates previous expert opinion. Second, the present study compared persistent and recovered children on individual measures of language skills rather than grouping all language skills together as has been done previously (e.g., Clark et al., 2017; Guitar & Conture, 2006). Therefore, present findings identified specific aspects of spoken language or speech production that may be more predictive of stuttering chronicity (e.g., performance on standardized measures of expressive and receptive language) than others (e.g., MLU). Third, present findings highlight the association between stuttering frequency and stuttering chronicity, which has received little prior attention. This revelation may be due to the inclusion of data from additional, and more recent studies (e.g., Roehl, 2018; Yairi & Ambrose 2005) than were included in previous reviews. For example, Clark et al. (2017) only cited literature as recent as 1999 from the Illinois Developmental Stuttering Project when reviewing longitudinal findings on stuttering frequency. Fourth, and lastly, given the quantitative nature of the present study, descriptive statistics of the clinical characteristics for the persistent and recovered children are reported.

Theoretical Connections

Present findings relate to the MDP model (Smith & Weber, 2017), which implicates multiple domains in the development of childhood stuttering. Specifically, the present results may be interpreted as evidence that there are genetic (e.g., male sex, family history) and speech-language (e.g., lower speech-sound accuracy and/or language) vulnerabilities that contribute to an individual's risk to exhibit persistent stuttering. The limited sample size relative to the evaluation of potential temperament/emotional contributions to persistence may underlie the non-significant between-group findings. Therefore, additional empirical studies will be necessary to establish a more comprehensive prospective for these variables (for overview of the potential role of these variables in stuttering, see Conture & Walden, 2012; Smith & Weber, 2017).

Ultimately, the continued empirical assessment of factors associated with risk for stuttering persistence will contribute to the development of related theoretical models, which, in turn, will likely lead to translational advancements. While keeping the aforementioned preliminary nature of results in mind, present findings can be used to inform clinical practice and future research studies. We discuss implications of the present findings relative to clinical practice and future research directions below.

Clinical Implications

Findings support and augment previous recommendations that SLPs should conduct comprehensive speech-language evaluations when working with young children who stutter and their families (e.g., Clark et al., 2017; Walsh et al., 2018). First, the present study reports descriptive statistics for persistent and recovered children, which

provide a reference point for the mean values and variability for each clinical characteristic. Second, by reporting effect sizes for each clinical characteristic, present findings might be used to identify a hierarchy of evaluation components to aid SLPs in the identification of factors that may be related to risk for stuttering persistence (as well as concomitant challenges that may need to be considered and/or addressed when planning treatment for these children) based on the evidence currently available. Table 3 summarizes the characteristics that might increase a child's risk for stuttering persistence along with related evaluation procedures based on the current level of evidence. Given the limited empirical evidence available, multi-faceted, comprehensive assessments that include characteristics found to have insufficient evidence and/or those not included in the present study are still warranted and provide valuable information. It should be underscored, data on clinical characteristics described as *insufficient* should not be interpreted as unimportant—but underexplored at present.

Table 3

Clinical characteristics related to stuttering chronicity (with established and insufficient evidence) organized by descending effect size

Clinical Characteristic	Evaluation Procedure
<i>Established Evidence</i>	
Known Family History of (Any) Stuttering	Caregiver Report
Male Sex	Caregiver Report
Older Age at Onset	Caregiver Report
Lower Speech-Sound Skills	Speech-Sound Test
Higher Rate of Stuttering-Like Disfluencies	Disfluency Sample
Lower Receptive Language	Global Receptive Language Test
Lower Expressive Language	Global Expressive Language Test
<i>Insufficient Evidence</i>	
Greater Negative/Positive Reactivity	Caregiver Questionnaire
Known Family History of Persistent Stuttering	Caregiver Report
Higher Rate of Individual Stuttering-Like Disfluency Types	Disfluency Sample
Expressive Vocabulary	Expressive Vocabulary Measure

Note. Established evidence characterized by statistically significant summary results. Insufficient evidence characterized by high heterogeneity across meta-analytic models, with at least one primary study reporting significant findings.

Whereas findings identify clinical characteristics that are related to stuttering persistence, additional research is needed to understand how SLPs can use these characteristics to assess a child's risk for persistence. For example, cumulative risk (i.e., the more risk factors exhibited, the greater the risk for persistence) has been proposed as important (e.g., Zebrowski, 1997), but it has not yet been empirically validated. Also, additional research is needed to determine whether there are clinical characteristics that are more valuable for predicting persistence than others and might be weighted more heavily and/or more critical to assess (refer to Table 3 for the ranking of present results by effect size).

Due to their high clinical utility, we encourage SLPs to contribute to practice-based evidence by collecting these data as part of their assessments of young children who stutter. By exploring the predictive value of these clinical characteristics in their own clients, SLPs could deepen their understanding of stuttering chronicity and, based on emerging evidence (Boaz, Hanney, Jones, & Soper, 2015), perhaps improve their own clinical procedures and outcomes as a result of collecting this data. For example, by comprehensively assessing the language skills of a child who stutters, a subtle, subclinical language delay is more likely to be identified and addressed in treatment and considered in regards to the child's risk for persistence. For SLPs interested in engaging in research more formally, the American Speech-Language-Hearing Association's Clinicians and Researchers Collaborating [CLARC] Initiative connects SLPs with researchers who have shared research interests. By capitalizing on the collaborative effort of clinicians and researchers, there is great potential to address a number of important research questions, including, but not limited to, the ideas described below, using large, representative datasets of young children who stutter.

Implications for Future Research

In addition to focusing on how to best assess risk of persistence, there are other directions for future research. First, additional cohort studies are necessary to assess whether other clinical characteristics may be related to stuttering persistence as well as to conduct moderator analyses. For example, only one study to date has explored nonword repetition tasks, which is thought to tap into phonological working memory as well as the speech-motor domain, a domain commonly thought to be important to the development

of stuttering. Additionally, the small number of eligible studies limited our ability to conduct moderator analyses when high heterogeneity was detected, which may have provided important insights on methodological characteristics that might have contributed to differences in findings across studies.

Second, group-level explorations, such as the present study, are limited in their ability to account for individual differences (i.e., heterogeneity) within groups. The reporting of data at the level of the individual would allow for the important exploration of whether certain profiles (e.g., a constellation of factors) might put a child at considerably greater risk for persistence than others. To explore profiles, large samples are necessary. Just as the sharing and assimilating of data across studies was essential to conduct the present study, we believe working towards the establishment of large, multisite databases is essential to the continued advancement of our understanding of stuttering development, assessment, and treatment. These databases would help to combat the low sample size issues that have plagued nearly all prospective cohort studies.

Further, research on these clinical characteristics might benefit from the development of a standard protocol for assessing young children. Developing a *flexible* standard protocol may facilitate the pooling of data and aligning of research and clinical practices. Based on present findings, we suggest that such a protocol should include collecting of speech samples for disfluency analysis, norm-referenced tests for speech-sound accuracy, expressive and receptive language, caregiver interviews that include collecting data on age at onset and family history of stuttering, and a caregiver report measure of temperament. By adopting similar assessment methods, but still collecting additional data of interest (e.g., physiological measures of reactivity, brain imaging, etc.),

clinicians and researchers could contribute valuable data central to furthering our understanding of stuttering persistence.

Last, it is important that researchers clearly report subject characteristics, methods and results in future studies. Based on present findings, our understanding of factors involved in stuttering persistence would benefit from clear reporting of attrition, methods to control for confounding variables, and samples that are representative of young children who stutter to reduce potential sources of bias. To facilitate future meta-analyses, data needed to calculate effect sizes should be reported as well (e.g., sample sizes for individual comparisons, means, standard deviations, etc.).

Caveats

Present findings were limited by our decisions when conducting the present meta-analysis. For example, more data would have been available had we allowed older children (i.e., above the age of six) at study entry. Similarly, had we adopted more restricted criteria for group classification (e.g., required direct evaluation of stuttering), less, but more homogenous data would have been available. Additionally, inherent to any meta-analysis, we acknowledge that despite our extensive search strategy, we may not have identified some relevant reports. Last, due to the small number of studies in some analyses, it is possible that false negatives are reported for some variables. Overall, we believe that the strengths of the present study outweigh the limitations. The present findings allowed us to evaluate clinical characteristics related to stuttering persistence based on the present literature base as well as identify potential areas worthy of future investigation.

CHAPTER V

CONCLUSION

The present study used meta-analytic methods to identify clinical characteristics that differentiate children who later persist versus those who recover from stuttering in early childhood. Male sex, a later onset of stuttering, a family history of stuttering, a higher rate of stuttering-like disfluency, lower speech-sound accuracy, and lower receptive and expressive language were found to be related to stuttering persistence. Insufficient evidence is available to support a relation between stuttering persistence and lower expressive vocabulary, greater negative and positive reactivity, and higher rates of individual stuttering-like disfluency types. Findings represent the *highest level of empirical evidence* for clinical risk factors associated with stuttering persistence to date and further our understanding of the nature of persistent stuttering. The inclusion of these clinical characteristics in the empirically-based initial evaluation of stuttering in young children will provide valuable diagnostic data that will aid speech-language pathologists in making treatment decisions. Future studies may consider investigating whether the presence of multiple risk factors confers increased risk for persistence (i.e., cumulative risk) as well as whether there are risk factors that confer greater risk than others.

REFERENCES

References marked with an asterisk indicate study data was included in the meta-analysis.

- Ambrose, N., Cox, N. J., & Yairi, E. (1997). The genetic basis of persistence and recovery in stuttering. *Journal of Speech, Language, and Hearing Research, 40*, 567–580.
- *Ambrose, N. G., Yairi, E., Loucks, T. M., Seery, C. H., & Throneburg, R. (2015). Relation of motor, linguistic and temperament factors in epidemiologic subtypes of persistent and recovered stuttering: Initial findings. *Journal of Fluency Disorders, 45*, 12–26.
- *Arenas, R. M., Walker, E. A., & Oleson, J. J. (2017). Developmental stuttering in children who are hard of hearing. *Language, Speech, and Hearing Services in Schools, 48*, 234–248.
- Bankson, N. W., & Bernthal, J. E. (1990). Bankson-Bernthal Test of Phonology. Austin, TX: Pro-Ed.
- Bernstein Ratner. (2018). Selecting treatments and monitoring outcomes: The circle of evidence-based practice and client-centered care in treating a preschool child who stutters. *Language, Speech, and Hearing Services in Schools, 49*, 13–22.
- Boaz, A., Hanney, S., Jones, T., & Soper, B. (2015). Does the engagement of clinicians and organisations in research improve healthcare performance: A three-stage review. *BMJ, 5*, 1-14.

- *Bostian, A. (2017). Disfluency characteristics of children aged 4-5 years associated with persistence and recovery of stuttering. (Unpublished master's thesis). Purdue University, West Lafayette, Indiana.
- Brosch, S., Hage, A., & Johannsen, H. S. (2001). Value of acoustic speech analysis for prognostic assessment of stuttering in children: Partial results of a prospective longitudinal study. *HNO*, *49*, 289–297.
- Brosch, S., Hage, A., Kolehne, P., & Johannsen, H. S. (1999). Stuttering children and the probability of remission—the role of cerebral dominance and speech production. *International Journal of Pediatric Otorhinolaryngology*, *47*, 71–76.
- *Buhr, A. (2007). *The roles of syntactic and prosodic components of grammar in early childhood stuttering* (Unpublished doctoral dissertation). University of Iowa, Iowa City, Iowa.
- Carrow-Woolfolk, E. (1999). Test for Auditory Comprehension of Language (3rd ed.). Circle Pines, MN: AGS.
- Chow, H. M., & Chang, S. E. (2017). White matter developmental trajectories associated with persistence and recovery of childhood stuttering. *Human brain mapping*, *38*, 3345-3359.
- Conture, E. G., & Walden, T. A. (2012). Dual diathesis-stressor model of stuttering. *Theoretical Issues of Fluency Disorders*, 94–127.
- Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002). Epidemiology of stuttering in the community across the entire life span. *Journal of Speech, Language, and Hearing Research*, *45*, 1097–1105.

- Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *Structured Photographic Expressive Language Test* (3rd ed.). DeKalb, IL: Janelle Publications.
- Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD, US: Paul H Brookes Publishing.
- Dunn, L. M., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test–III*. Circle Pines, MN: AGS.
- *Dworzynski, Remington, Rijdsdijk, Howell, & Plomin. (2007). Genetic etiology in cases of recovered and persistent stuttering in an unselected, longitudinal sample of young twins. *American Journal of Speech-Language Pathology, 16*, 169–178.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629–634.
- *Erdemir, A., Walden, T. A., Jefferson, C. M., Choi, D., & Jones, R. M. (2018). The effect of emotion on articulation rate in persistence and recovery of childhood stuttering. *Journal of Fluency Disorders, 56*, 1–17.
- Fenson, L., Dale, P. S., & Reznick, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.
- *Franken, M.-C. J. P., Koenraads, S. P. C., Holtmaat, C. E. M., & van der Schroeff, M. P. (2018). Recovery from stuttering in preschool-age children: 9-year outcomes in a clinical population. *Journal of Fluency Disorders, 58*, 35–46.
- Garnett, E. O., Chow, H. M., Nieto-Castañón, A., Tourville, J. A., Guenther, F. H., & Chang, S.-E. (2018). Anomalous morphology in left hemisphere motor and premotor cortex of children who stutter. *Brain, 141*, 2670–2684.

- Goldman, R., & Fristoe, M. (2000). Goldman-Fristoe Test of Articulation—Second Edition. Circle Pines, MN: AGS.
- Gordon, P. A., & Luper, H. L. (1992). The early identification of beginning stuttering I: Protocols. *American Journal of Speech-Language Pathology, 1*, 43–53.
- Guitar, B., & Conture, E. G. (Eds.). (2006). *The child who stutters: To the pediatrician* (Third edition). Memphis, TN: Stuttering Foundation of America.
- Hage, A. (2001). Cognitive and Linguistic Abilities in Young Children: Are They Able to Predict the Further Development of Stuttering. *Sprache Stimme Gehör, 25*, 20–24.
- Hall, K. D., Amir, O., & Yairi, E. (1999). A longitudinal investigation of speaking rate in preschool children who stutter. *Journal of Speech, Language, and Hearing Research, 42*, 1367–1377.
- Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine, 158*, 280–289.
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*, 490–499.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*, 557–560.
- Hilger, A. I., Zelaznik, H., & Smith, A. (2016). Evidence that bimanual motor timing performance is not a significant factor in developmental stuttering. *Journal of Speech, Language and Hearing Research, 59*, 674-685.
- Hodson, B. W. (1986). *The Assessment of Phonological Processes-Revised*. Austin, Danville, IL: Interstate.

- *Hollister, J., Van Horne, A. O., & Zebrowski, P. (2017). The relationship between grammatical development and disfluencies in preschool children who stutter and those who recover. *American Journal of Speech-Language Pathology, 26*, 44–56.
- Hresko, W. P., Reid, D. K., & Hammill, D. D. (1999). *Test of Early Language Development—Third Edition*. Austin, TX: Pro-Ed.
- Jameson, A. M. (1955). Stammering in children - some factors in the prognosis. *Speech, 19*, 60–67.
- Johnson, W. (1959). *The onset of stuttering: Research findings and implications*. University of Minnesota Press.
- *Kefalianos, E., Onslow, M., Packman, A., Vogel, A., Pezic, A., Mensah, F., ... Reilly, S. (2017). The history of stuttering by 7 years of age: Follow-up of a prospective community cohort. *Journal of Speech, Language, and Hearing Research, 60*, 2828–2839.
- Kelman, E., & Nicholas, A. (2008). *Practical Intervention for Early Childhood Stammering: Palin PCI Approach*. London, United Kingdom: Speechmark Publishing Ltd.
- *Kloth, S. A. M., Kraaimaat, F. W., Janssen, P., & Brutten, G. J. (1999). Persistence and remission of incipient stuttering among high-risk children. *Journal of Fluency Disorders, 24*, 253–265.
- Lee, L. L., & Canter, S. M. (1971). Developmental Sentence Scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*, 315–340.

- *Leech, K., Bernstein Ratner, N., Brown, B., & Weber, C. (2017). Preliminary evidence that growth in productive language differentiates childhood stuttering persistence and recovery. *Journal of Speech, Language, and Hearing Research, 60*, 3097-3109.
- Mahurin-Smith, J., & Ambrose, N. G. (2013). Breastfeeding may protect against persistent stuttering. *Journal of Communication Disorders, 46*, 351–360.
- Månsson, H. (2000). Childhood stuttering: Incidence and development. *Journal of Fluency Disorders, 25*, 47–57.
- Mansson, H. (2005). Stammens kompleksitet og diversitet. *Dansk Audiologopædi, 41*, 13–33.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology, 62*, 1006–1012.
- Nippold, M. A. (2018). Stuttering in preschool children: Direct versus indirect treatment. *Language, Speech, and Hearing Services in Schools, 49*, 4–12.
- Onslow, M., Andrews, C., & Costa, L. (1990). Parental severity scaling of early stuttered speech: Four case studies. *Australian Journal of Human Communication Disorders, 18*, 47–61.
- Onslow, M., & Packman, A. (1999). The Lidcombe Programme and natural recovery: Potential choices of initial Management strategies for early stuttering. *Advances in Speech Language Pathology, 1*, 113–121.

- Paden, E. P., Ambrose, N. G., & Yairi, E. (2002). Phonological progress during the first 2 years of stuttering. *Journal of Speech, Language, and Hearing Research, 45*, 256–267.
- Paden, E. P., & Yairi, E. (1996). Phonological characteristics of children whose stuttering persisted or recovered. *Journal of Speech, Language, and Hearing Research, 39*, 981–990.
- Paden, E. P., Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering II: Initial status of phonological abilities. *Journal of Speech Language and Hearing Research, 42*, 1113-1124.
- Reilly, S., Onslow, M., Packman, A., Wake, M., Bavin, E. L., Prior, M., ... Ukoumunne, O. C. (2009). Predicting stuttering onset by the age of 3 years: A prospective, community cohort study. *PEDIATRICS, 123*, 270–277.
- Reynell, J. (1983). *Reynell Developmental Language Scale Manual, Revised edition*. Windsor: Nfer-Nelson Publishing Company Ltd.
- Riley, G. D. (1972). A Stuttering Severity Instrument for Children and Adults. *Journal of Speech and Hearing Disorders, 37*, 314–322.
- *Roehl, L. (2018). *Neural Systems Supporting Nonlinguistic Auditory Processing in Young Children Who Persist and Recover from Stuttering* (Master's Thesis). Michigan State University, Lansing, Michigan.
- *Rommel, D., Hage, A., Kalebne, P., & Johannsen, H. S. (2001). Development, maintenance, and recovery of childhood stuttering: Prospective longitudinal data three years after first contact. In K.L. Baker, L. Rustin, & F. Cook (Eds.),

- Proceedings of the Fifth Oxford Dysfluency Conference*. Windsor: Chappel Garden.
- Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly*, *40*, 21–39.
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The children’s behavior questionnaire. *Child Development*, *72*, 1394–1408.
- *Ryan, B. P. (2001). A longitudinal study of articulation, language, rate, and fluency of 22 preschool children who stutter. *Journal of Fluency Disorders*, *26*, 107–127.
- Sawyer, J. (2005). *Characteristics of disfluency clusters in stuttering and normally fluent preschool children: A longitudinal study* (Doctoral Dissertation). University of Illinois, Urbana Champagne, Illinois.
- Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, *11*, 1–22.
- Shimada, M., Toyomura, A., Fujii, T., & Minami, T. (2018). Children who stutter at 3 years of age: A community-based study. *Journal of Fluency Disorders*, *56*, 45–54.
- *Singer, C. M., Walden, T. A., & Jones, R. M. (2019a). *Attention, speech-language dissociations and stuttering chronicity*. (Unpublished Manuscript).
- Singer, C. M., Walden, T. A., & Jones, R. M. (2019b). Differences in the relation between temperament and vocabulary based on children’s stuttering trajectories. *Journal of Communication Disorders*, *78*, 57–68.

- Smith, A., & Weber, C. (2017). How stuttering develops: The Multifactorial Dynamic Pathways Theory. *Journal of Speech, Language, and Hearing Research, 60*, 2483–2505.
- *Spencer, C., & Weber-Fox, C. (2014). Preschool speech articulation and nonword repetition abilities may help predict eventual recovery or persistence of stuttering. *Journal of Fluency Disorders, 41*, 32–46.
- Subramanian, A., Yairi, E., & Amir, O. (2003). Second formant transitions in fluent speech of persistent and recovered preschool children who stutter. *Journal of Communication Disorders, 36*, 59–75.
- Throneburg, R. N., & Yairi, E. (2001). Durational, proportionate, and absolute frequency characteristics of disfluencies. *Journal of Speech, Language, and Hearing Research, 44*, 38–51.
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*, 375–393.
- Tumanova, V., Conture, E. G., Lambert, E. W., & Walden, T. A. (2014). Speech disfluencies of preschool-age children who do and do not stutter. *Journal of Communication Disorders, 49*, 25–41.
- *Walsh, B. (2019). Personal communication.
- *Walsh, B., Usler, E., Bostian, A., Mohan, R., Gerwin, K., Brown, B., ... Smith, A. (2018). What are predictors for persistence in childhood stuttering? *Seminars in Speech and Language, 39*, 299–312.

- Watkins, R. V., & Yairi, E. (1997). Language production abilities of children whose stuttering persisted or recovered. *Journal of Speech, Language, and Hearing Research, 40*, 385–399.
- *Watkins, R. V., Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering III: Initial status of expressive language abilities. *Journal of Speech, Language, and Hearing Research, 42*, 1125–1135.
- Williams, K. (1997). Expressive Vocabulary Test—Second Edition. Circle Pines, MN: AGS
- Yairi, E., & Ambrose, N. (1992). Onset of stuttering in preschool children. *Journal of Speech, Language, and Hearing Research, 35*, 782–788.
- Yairi, E., & Ambrose, N. (1999). Early childhood stuttering I: Persistence and recovery rates. *Journal of Speech, Language, and Hearing Research, 42*, 1097–1112.
- Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders, 38*, 66–87.
- *Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering for clinicians by clinicians*. Austin, Tex.: PRO-ED.
- Yairi, E., Ambrose, N. G., Paden, E. P., & Throneburg, R. N. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering. *Journal of Communication Disorders, 29*, 51–77.
- Zebrowski, P. M. (1997). Assisting young children who stutter and their families: Defining the role of the speech-language pathologist. *American Journal of Speech-Language Pathology, 6*, 19–28.

*Zengin-Bolatkale, H., Conture, E. G., Walden, T. A., & Jones, R. M. (2018).

Sympathetic arousal as a marker of chronicity in childhood stuttering.

Developmental Neuropsychology, 43, 135–151.

Zimmerman, I., Steiner, V., & Pond, R. (1979). Preschool language scale. The

Psychological Corporation, Harcourt Brace Jovanovich, Inc.

Appendix A

Report Characteristics by Study

Report	Type	Additional Eligibility Criteria ¹	Persisting	Recovered	Follow-Up Length	Confounding	Clinical Characteristic Reported
California - Long Beach							
Ryan (2001)	PR		7 (4M) 50 mo. (11)	15 (10 M)	2-10 years	NR	S, [SS], [RV], [GL], [SF]
Early Language in Victoria Study							
Kefalianos et al. (2017)	PR		36 (24 M) 48 mo.	67 (37 M) 48 mo.	at least 3 years	adjusted values reported	S, EV, FH, T
Erasmus University Medical Centre Study							
Franken et al. (2018)	PR		4 (4 M) 52.25 mo. (7.0)	11 (6 M) 43.36 mo. (8.6)	at least 9 years	N/A	S, A, FH, FH-P, FH-R, PSS, CSS, SF
Illinois Stuttering Research Project							
Watkins, Yairi & Ambrose (1999)	PR		22 (18 M) 43 mo.	62 (40 M) 39 mo.	at least 4 years	data stratified by age	S, MLU, NDW*, NTW*, DSS
Yairi & Ambrose (2005)	BK	within 12 mo. of onset	19 (15 M) 43.16 mo. (9.52)	70 (49 M) 36.87 mo. (7.66)	at least 4 years	multiple methods reported	S, SS ² , EL, EV, FH, FH-P, [PSS], [CSS], SF
Netherlands- High Risk							
Kloth, Kraaimaat, Janssen, & Brtten (1999)	PR		7 (5 M) 40 mo. (6.45)	16 (7 M) 36 mo. (9.8)	at least 4 years	NR	S, EL, RL, MLU
Outcomes of Children with Hearing Loss							
Arenas, Walker, & Oleson (2017)	PR		7	11	at least 2 years or until 8 years old	NR	S, SS, EL, RL

¹ Criteria adopted in addition to the study-wide criteria.

² Reported for 19 Persistent children and 65 Recovered children.

Purdue Stuttering Project

Bostian (2017)	UP	4 to 5 years old at entry	19 (14 M) 58.05 mo. (1.71)	29 (22 M) 53.58 mo. (1.04)		N/A	S, SF, OD, PW, SW, DP, MR* , CSS
Leech, Bernstein Ratner, Brown, & Weber (2017)	PR		28 (15 M) 62 mo.	22 (11 M) 57 mo.	2 years	age and SES not statistically different	S, A, SS, RL, EL, NRT*, IPSYN ,
Roehl (2018)	UP		16 (12 M) 57.72 mo. (1.7)	26 (17 M) 54.84 mo. (1.3)	2-5 years	NR	S, A, EL, RL, SF, CSS, PSS
Spencer & Weber-Fox (2014)	PR		19 (15 M) 57.11 mo. (6.71)	21 (14 M) 53.33 mo. (5.36)	12-48 months	age, nonverbal reasoning, SES matching	S, SS , EL, RL, NRT*
Walsh et al. (2018) ³	PR		31 (26 M)	32 (21 M)			S, FH⁴, FH-P⁴, FH-R⁴
Subtypes and Risk Factors							
Ambrose, Yairi, Loucks, Seery, & Throneburg (2015)	PR		19 (12 M) 43.63 mo. (11.96)	39 (27 M) 38.18 mo. (7.57)	up to 5 years	NR	S, A, EL, RL, EV, EL , [SS], [MLU], CSS, PSS, SF, PW, SW, DP, RU* , T
Buhr (2007)	UP	language skills within normal limits	4 49 mo.	4 50 mo.	2 years	NR	GL*, DSS, MLU, SF, OD
Hollister, Van Horne, & Zebrowski (2017)	PR	excluded children > 43 months or if speech sample < 100 utterances	8 37.1 mo. (4.6)	5 35.2 mo. (5.8)	2 years	age not statistically different	EL, RL, EV, RV, SF, CSS, IPSYN, MLU
Twins Early Developmental Study							
Dworzynski, Remington, Rijdsdijk, Howell & Plomin (2007)	PR		135 (100 M)	950 (521 M)		N/A	S
The Ulm Study							
Rommel, Hage, Kalehne, & Johannsen (1999)	UP		19 (16 M) 62.4 mo.	46 (33 M) 58.8 mo.	3 years	NR	S, [A], FH, [RV], [L], [MLU], [NDW] ,

³ Additional data for calculated effect sizes collected through personal communication via email with the first author (Walsh, 2019)

⁴ Reported for 17 persistent children and 28 recovered children

Vanderbilt Developmental Stuttering Project

Erdemir, Walden, Jefferson, Choi, & Jones (2018)	PR		10(9 M) 46.9 mo. (4.5)	10 (9 M) 46.1 mo. (6.9)	2-2.5 years	sex, age, and language matched	S, A , SF, CSS
Singer, Walden, & Jones (2019a)	UP	under 5;0 years. at entry	10 (9 M) 46.6 mo. (4.5)	32 (23 M) 45.0 mo. (6.8)	at least 2 years	age and SES not statistically different	S, CSS, SF, SS, EL, RL, EV, RL
Zengin-Bolatkale, Conture, Walden, & Jones (2018)	PR		9 (8 M) 45.11 mo. (6.23)	23 (17 M) 47.11 mo. (6.84)	16-32 months	age not statistically different	S, CSS, SF, SS, EL, RL, EV, RV, CRS*, T

Note. Bold indicates data were extracted for the present study; Brackets indicate data could not be used to estimate an effect size; Asterisk indicates data not reported in another study; PR = Published Report; BK = Book; UP = unpublished; NR = Not Reported, S = Sex, A = Age at onset; FH = Family History of Stuttering; FH-P = Family History of Persistent Stuttering; FH-R = Family History of Recovered Stuttering; CSS = Clinician-Rated Stuttering Severity; PSS = Parent-Rated Stuttering Severity; SF = Stuttered-Like Disfluency Frequency, NSF = Non-stuttered-like disfluency Frequency; PW = Part-Word Repetition Frequency; SW= Single-Word Repetition Frequency; DP = Dysrhythmic Phonation Frequency; RU = Repetition Units (mean), MR = Max Repetition Units, SS = Speech-Sound Accuracy; GL = Global Language; EL = Expressive Language; RL = Receptive Language; EV = Expressive Vocabulary; RV = Receptive Vocabulary; IPSYN = Index of Productive Syntax (IPSYN; Scarborough, 1990); MLU = Mean Length Utterance; DSS = Developmental Sentence Score (Lee & Canter, 1971); CRS = Child Response to Stuttering; T = Temperament.

Appendix B

Speech Sample Characteristics by Study

Study	Speech Sample Characteristics			
	Duration	Word/Syllable Count	Communication Partner	Activity Examples
Erasmus Medical Center	30 min x 2	1,000 syllables	Caregiver	NR
Illinois Stuttering Research Project	NR	1,000 words	Primary Caregiver	Play-doh
Netherlands – High Risk	30 min (10 min coded)	NR	Mother	NR
Purdue Stuttering Project	12 min x 2	750-1,000 words	Caretaker; SLP	Clay
Subtypes and Risk Factors in Childhood Stuttering	20 min x 2	1,000+ syllables	Primary Caregiver; SLP	play-doh
Vanderbilt Developmental Stuttering Project	NR	300 words	Clinician	Farm set

Note. NR = Not Reported; SLP = Speech-Language Pathologist.